

La biométrie des caractères discontinus d'après le revêtement appendiculaire des Collemboles Symphyléones

I. Sur quelques concepts nouveaux dans l'analyse numérique de la chétotaxie

par Pierre NAYROLLES

Résumé. — Nous montrons que l'étude de la chétotaxie peut se réaliser à travers un cadre conceptuel faisant référence à la théorie des ensembles. Si nous considérons l'ensemble des soies observées sur plusieurs individus de la même espèce, à un même stade de développement et pour un même sexe, cet ensemble de soies peut être subdivisé jusqu'à des éléments qui ne sont plus décomposables. Dans le cas où les soies sont idionymiques, chacun de ces éléments correspond à une soie. La variabilité des soies idionymiques est ensuite étudiée ; nous proposons une fonction de description, appelée occurrence, pour rendre compte du taux de présence d'une soie. Les soies peuvent présenter des liaisons statistiques entre elles. Ce phénomène est notamment illustré par la compensation qui correspond à une variabilité observée plus petite que celle qui devrait être dans le cas où chaque soie est indépendante des autres.

Mots-clés. — Biométrie, caractères discontinus, chétotaxie, Collemboles, Symphyléones.

Abstract. — We show that the chaetotaxic study can be achieved with the help of the set theory as a conceptual framework. Consider any setal set observed into several specimens belonging to the same species, same instar and same sex, this set can be shared up to indivisible components. If all the setae are idionymic, each component amounts to a seta. Afterward idionymic setal variability is studied, for that, we propose a function of description giving the setal presence ratio, it is called occurrence. Statistic relationships among setae are put forwards. Particularly, a relevant fact of this phenomenon lies in the compensation, which occurs when the observed variability is less than the variability that one should observe if each seta were independent from the others.

Key words. — Biometry, discontinuous characters, chaetotaxy, Collembola, Symphyleona.

P. NAYROLLES, *Laboratoire de Zoologie, Écobiologie des Arthropodes édaphiques, UPR CNRS 90 14, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex.*

INTRODUCTION

La variabilité des phanères est un fait constaté depuis déjà longtemps. Sur le groupe des Acariens F. GRANDJEAN s'est intéressé aux problèmes qui en découlent. Cet auteur définit un écart (GRANDJEAN, 1939, 1942, 1949a) comme à la fois l'existence et la fréquence de la plus rare des deux modalités : présence (que nous notons P) et absence (que nous notons A). La proportion souvent importante d'exemplaires dissymétriques pour un organe pair variable peut

s'expliquer en considérant que la distribution des probabilités des divers types d'individus, à savoir : PP, PA et AA suit une loi binomiale. Cela correspond à l'asymétrie au sens de GRANDJEAN (1939). Chez les Collemboles Symphyléones nous avons très souvent rencontré ce type de variabilité que GRANDJEAN (1949a) distingue de ce qu'il nomme « mutation », cette dernière entraînant l'absence d'individus dissymétriques dans la population (les animaux s'observent sous deux formes : soit PP, soit AA).

Dans ce travail plusieurs problèmes seront abordés : comment isoler et caractériser ces « objets de taxonomie » que constituent les phanères ? Au sein d'une espèce et même d'un stade du développement, le systématicien comparera plusieurs individus ; pour chacun de ces « objets » que retiendra-t-il de ces comparaisons ? Enfin nous montrerons que ces « objets » établissent parfois entre eux des relations ; ils peuvent présenter une « attitude » collective.

Nous rappellerons quelques aspects de la chétotaxie appendiculaire des Collemboles Symphyléones, pour la plupart tirés de notre étude de l'antenne (NAYROLLES, 1991a). Ces rappels seront utilisés comme exemples dans nos propos.

Une grande partie du travail que nous présentons ici a déjà été exposée au cours d'un travail de thèse (NAYROLLES, 1991b). Lors de cette étude, nous avons d'abord recherché des objets descripteurs permettant de distinguer à la fois les espèces récoltées et leurs stades de développement (ce fut l'étude de la chétotaxie appendiculaire) ; nous avons ensuite été amené à développer des considérations théoriques visant à établir d'une part les caractéristiques du revêtement, d'autre part le traitement d'un tableau tridimensionnel espèces/stades/caractères ; enfin, ces vues théoriques ont trouvé une application pratique dans l'analyse ontophylogénétique du matériel étudié (espèces appartenant aux familles des Bourletiellidae et des Sminthuridae, récoltées dans le sud-ouest de la France et le nord-est de l'Espagne). Le sujet du présent article est donc limité à la partie théorique ayant pour objectif de fonder un outil conceptuel destiné à l'étude des caractéristiques du revêtement.

I. QUELQUES NOTIONS PRÉALABLES SUR LA CHÉTOTAXIE

Les explications que nous allons fournir feront référence à la nomenclature que nous avons donnée de la chétotaxie antennaire (NAYROLLES, 1991a). BRETFELD (1990) a aussi donné une étude exhaustive de la chétotaxie appendiculaire de quatre espèces. Cependant, ce travail est limité à la seule famille des Bourletiellidae, or la chétotaxie antennaire est très différente selon les familles de Symphyléones ; en conséquence le travail de BRETFELD, bien qu'irréprochable dans l'exactitude de ses descriptions, ne peut fournir une nomenclature générale de la chétotaxie antennaire des Symphyléones.

En comparant la chétotaxie d'une région du corps entre plusieurs individus pris au même stade, on peut observer soit que les soies présentent une individualité propre (elles peuvent alors être dénommées), soit qu'elles varient dans leur position relative (dans ce cas là, généralement, le nombre de soies varie également). Le premier type correspond à ce que GRANDJEAN (1949b) désigne par le terme d'*idionymie*.

Nous avons défini (NAYROLLES, 1991a, 1991b, 1992) un ensemble chétotaxique comme la collection des soies situées sur une zone morphologique. Un ensemble chétotaxique s'applique à une espèce donnée (et plus précisément à une population d'une espèce), pour un stade de

développement et éventuellement pour un sexe précis. Il est dit **défini** si les soies qui le composent sont toutes idionymiques ; il est dit **structuré** si la répartition topologique des soies présente une organisation.

Les phanères des appendices des Symphypléones s'organisent en effet en rangées longitudinales, les génératrices, désignées par la lettre G. Sur le quatrième article antennaire, existent en outre des intergénératrices, désignées par la lettre H. On parlera de soies G ou de soies H selon leur appartenance à l'une ou à l'autre de ces deux catégories. On compte huit génératrices : Ge, Gae, Ga, Gai, Gi, Gpi, Gp et Gpe qui sont respectivement en position externe, antéro-externe, antérieure, antéro-interne, interne, postéro-interne, postérieure et postéro-externe ; et huit intergénératrices : Heae, Haea, Haai, Haii, Hipi, Hpip, Hppe et Hpee. L'intergénératrice Heae est située entre les génératrices Ge et Gae, Haea entre Gae et Ga, etc.

Outre cette structure longitudinale, on peut distinguer une organisation transversale qui découpe le quatrième article antennaire en trois tronçons : A, M et B (A pour apical, M pour médian et B pour basal). Chaque tronçon peut à son tour être découpé en zones (pour le tronçon B), ou en verticilles (pour les tronçons A et M). Les zones du tronçon B sont écrites BA, BM et BB (ou encore BAM si BA n'est pas individualisée de BM, BMB si BM n'est pas individualisée de BB). Les verticilles de M sont numérotés de l'apex vers la base : M1, M2, M3, etc. Nous avons adopté un tel sens de numérotation car les structures les plus conservées entre taxons sont toujours distales (ceci est même la règle générale pour la chétotaxie appendiculaire des Symphypléones, voir NAYROLLES, 1992).

La figure 1 illustre la répartition des phanères sur le quatrième article antennaire de *Bourletiella hortensti* (Fitch, 1863) et la schématisation de la structure chétotaxique qu'on peut y associer. Le tronçon A est divisé en 3 verticilles : AI, AII et AIII. Le tronçon M comprend 5 verticilles : M1, M2, M3, M4 et M5. Dans le tronçon B nous pouvons distinguer 3 zones : BA, BM et BB. Par rapport à l'ensemble de la chétotaxie du quatrième article antennaire, seule la zone BM présente une chétotaxie non définie. Outre son revêtement, le quatrième article antennaire est caractérisé par sa morphologie qui est ici subsegmentée. On compte un subsegment pour le tronçon A, un pour chacun des verticilles de M et un pour le tronçon B. La figure 2 présente quatre exemples de revêtement du tronçon B ; elle illustre la chétotaxie structurée non définie de BM. BRETTFELD (1990) a noté que la portion moyenne et apicale de BM, qu'il désigne par la lettre T, présente chez les espèces de Bourletiellidae une chétotaxie très fluctuante ; comparativement les soies basales de BM ont une position plus stable.

Pour nommer une soie, nous combinons la coordonnée du verticille ou du tronçon à celui de la génératrice ou de l'intergénératrice. Ce symbole pourra être précédé de deux lettres entre parenthèses qui indiquent le segment antennaire en question : (AQ) pour le quatrième article antennaire. Exemple : (AQ)M2i est la soie du quatrième article antennaire sur le verticille M2 et sur la génératrice Gi.

La plupart des exemples que nous donnerons concernent les soies du quatrième article antennaire des Symphypléones. Dans ce travail (comme dans les précédents), le mot « stade » est employé au sens de « stase » de GRANDJEAN (1938), le premier étant d'usage général parmi les collembologistes. Avant de modifier le vocabulaire, sans doute vaut-il mieux attendre des études sur le développement ontogénétique des Arthropodés... Cependant, ANDRÉ (1986, 1988, 1989) a discuté de la notion de stase chez les Arthropodés et a proposé d'étendre l'usage de ce mot au cas des Collembolés, ce qui pose certains problèmes dont la discussion dépasse le cadre de ce travail.

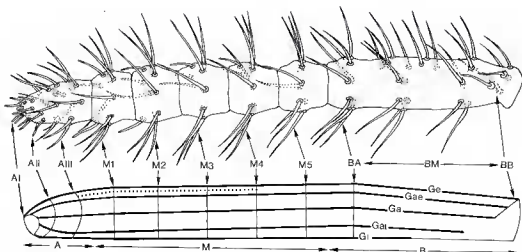


FIG. 1. — Quatrième article antennaire de *Bourlettiella hortensis* (d'après NAYROLLES, 1992b). La figure du haut est un dessin du revêtement en vue antérieure. La figure du bas donne une représentation schématique de la chétotaxie de la face antérieure. Les génératrices sont dessinées en trait épais, l'intergénératrice Heae en pointillé, les verticilles en trait fin.

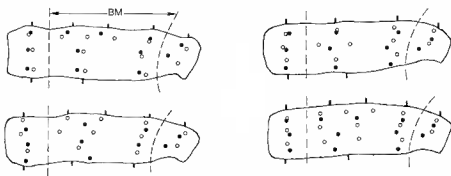


FIG. 2. — Quatre exemples de chétotaxie du tronçon B du quatrième article antennaire de *Bourlettiella hortensis* (d'après NAYROLLES, 1992b). Des tirets délimitent la zone BM. Les soies des génératrices Ge (en haut) et Gi (en bas) sont représentées par des petits traits. Les soies de la face antérieure sont figurées par des ronds pleins, celles de la face postérieure par des ronds vides.

II. LES ENTITÉS ET LES VARIABLES CHÉTOTAXIQUES

Soit un ensemble chétotaxique d'une espèce donnée, pour un sexe donné, à un stade donné; si cet ensemble est défini ou structuré, alors nous pouvons le « décomposer » en plusieurs sous-ensembles. Par exemple, sur la zone BM du quatrième article antennaire (nous considérons le cas général où la chétotaxie de BM n'est pas définie) nous pouvons distinguer :

- le sous-ensemble des soies des génératrices Ge, Gae, Ga et Gai de BM ;
- le sous-ensemble des soies des génératrices Gi, Gpi, Gp et Gpe de BM.

Au sens de la théorie des ensembles, il est entendu qu'un sous-ensemble chétotaxique est lui-même un ensemble chétotaxique.

Nous avons vu que sur l'antenne la structure du revêtement est établie selon deux composantes qui « s'entrecroisent à angle droit » :

- une organisation longitudinale qui découpe l'antenne en génératrices et intergénératrices ;
- une organisation transversale qui découpe l'antenne en articles, tronçons, zones et verticilles.

L'ensemble des soies de Ge, Gae, Ga et Gai de la zone BM peut ainsi être considéré comme l'intersection (symbole $\cap$) de l'ensemble des génératrices Ge, Gae, Ga et Gai de l'antenne avec l'ensemble constitué par la zone BM du quatrième article antennaire; et nous pouvons donc l'écrire sous la forme : $\{Ge, Gae, Ga, Gai\} \cap \{(AQ)BM\}$, ou plus simplement $\{Ge, Gae, Ga, Gai\} \cap \{BM\}$. En utilisant le symbole de la réunion ($\cup$), l'expression précédente peut encore s'écrire : $\{Ge \cup Gae \cup Ga \cup Gai\} \cap \{BM\}$. Évidemment nous pouvons remarquer que lorsque aucune soie H n'est présente sur la zone BM (cas le plus commun), nous avons : $\{Ge, Gae, Ga, Gai, Gi, Gpi, Gp, Gpe\} \cap \{BM\} = \{BM\}$.

$\{BM\}$ est l'ensemble chétotaxique déterminé sur une espèce, pour un stade, et même pour un sexe donné, qui correspond à la zone BM. Cette dernière a été établie sur l'ensemble du groupe des Symphyléones; sa nature a une valeur générale. En effet, son existence n'a pu être précisée qu'après avoir distingué des ensembles chétotaxiques $\{BM\}$ chez plusieurs espèces, à plusieurs stades. Ainsi, $\{BM\}$ et BM ne sont pas synonymes. Nous pourrions cependant les confondre dans leurs écritures, en écrivant plus simplement BM au lieu de $\{BM\}$. Les accolades seront utilisées pour des ensembles dans lesquels seront énumérés plusieurs éléments.

Réunissons sous le terme de **contexture** les notions d'idionymie et de structure. Pour tout ensemble chétotaxique structuré non défini (comme BM), il existe une partition telle que chaque sous-ensemble obtenu ne peut plus être divisé à nouveau. Ce « découpage » est maximal et les sous-ensembles qu'il engendre ne possèdent plus de contexture propre. Nous dirons que dans cette opération nous avons épuisé par partition la contexture de l'ensemble chétotaxique initial.

Par exemple, dans l'ensemble $\{Ge, Gae, Ga, Gai\} \cap BM$, nous pouvons distinguer les deux sous-ensembles suivants : $Ge \cap BM$ et $\{Gae, Ga, Gai\} \cap BM$. Nous ne pouvons plus découper $Ge \cap BM$; ce qui n'est pas le cas de $\{Gae, Ga, Gai\} \cap BM$. Le premier de ces deux sous-ensembles ne présente plus de partition possible car les soies qu'il porte ne sont pas véritablement définies individuellement, elles sont variables simultanément dans leurs positions

et dans leur quantité : elles ne sont pas idionymiques. Un tel ensemble chétotaxique indivisible sera appelé **entité chétotaxique**.

La même opération dans un ensemble chétotaxique défini (qu'il soit structuré ou non) amène à des entités chétotaxiques réduites chacune à une unique soie. Nous dirons alors que nous avons à faire à des **entités chétotaxiques singulières**. Dans le cas précédent, où une entité chétotaxique contient généralement plusieurs soies, nous sommes en présence d'**entités chétotaxiques plurielles**.

Alors que jusqu'ici nous nous situons dans un cadre intraspécifique et même pour un stade donné du développement, nous allons maintenant aborder les comparaisons de nature ontophylogénétique. Il semble évident que ces dernières se ramènent à des problèmes d'homologie entre entités chétotaxiques.

Pour le cas des entités chétotaxiques plurielles, nous pouvons relever le problème suivant : la structure du revêtement d'une région peut correspondre à un ensemble chétotaxique non défini (exemple : BM) dans la plupart des taxons d'un phylum, avec cependant des espèces ou des groupes d'espèces pour lesquels ce même ensemble chétotaxique est défini (exemples : BM de *Lipothrix lubbocki* (Tullberg, 1872), BMB des Dicyrtomidae). Pour ces espèces, nous garderons généralement comme entités chétotaxiques celles qui ont été définies sur l'ensemble du groupe.

Ainsi pour les espèces de Dicyrtomidae observées nous considérons que chaque génératrice de BMB constitue une entité chétotaxique, y compris $Ge \cap BMB$ qui contient deux soies fixes. Néanmoins, en vue d'une étude plus précise sur la seule famille des Dicyrtomidae, nous pourrions diviser $Ge \cap BMB$ en ses deux soies constitutives, chacune considérée comme une entité chétotaxique singulière : cette partition là serait maximale. Faire intervenir des comparaisons ontophylogénétiques pour le but d'obtenir des entités chétotaxiques « globales » rend donc légèrement fluctuant le niveau de la partition maximale.

Un cas particulier et paradoxal est celui du quatrième article antennaire de *Sminthurus viridis* (Linné, 1758). Chez cette espèce, l'antenne existe sous deux formes : l'une ayant 14 verticilles sur le tronçon M, l'autre en ayant 15. Les deux articles antennaires d'un même individu ont le même nombre de verticilles sur M. Pour moitié, les exemplaires examinés (50 au total) ont des antennes avec 14 verticilles sur M, les autres ont des antennes à 15 verticilles sur M. Nous avons pu montrer que le tronçon M est lié par interrelation à la zone BM (statistiquement le nombre de soies sur celle-ci dépend du nombre de verticilles sur celui-là). Il conviendrait donc de considérer la réunion de ces deux ensembles : $BM \cup M$. En toute rigueur, les entités chétotaxiques ne peuvent donc être définies que de manière indépendante sur les deux types distincts d'individus. Il sera nécessaire de proposer des descriptions rendant compte des caractéristiques communes du tronçon M entre les animaux à 14 subsegments intermédiaires et ceux qui en ont 15.

Une autre attitude consisterait à vouloir traiter d'emblée l'espèce *S. viridis* comme un tout, et de définir les entités chétotaxiques sur la portion s'étendant de M1 jusqu'à BM ; ces entités chétotaxiques apparaissant alors comme étant les génératrices et les intergénératrices de la région antennaire considérée. Manifestement, appliquer un tel raisonnement ne pourrait rendre compte des observations réalisées. En vérité il existe une complémentarité, et non pas une opposition, entre les deux façons de voir ce problème : dans la réalisation des observations le premier point de vue prévaut, la recherche de caractéristiques communes aux deux types d'individus qui s'en suit se rapporte plutôt à la deuxième conception.

Pour une espèce donnée et pour un type d'individu donné (il faut tenir compte du sexe, du stade de développement et éventuellement du type morphologique comme chez *S. viridis*), nous pouvons construire la matrice des observations chétotaxiques, dont les colonnes sont les entités chétotaxiques et les lignes les différentes observations.

Ainsi définies, les entités chétotaxiques apparaissent comme des vecteurs sur lesquels nous pouvons effectuer des combinaisons linéaires (multiplication par un scalaire, addition vectorielle), ou non (par exemple formation d'un vecteur par multiplication des coordonnées de deux vecteurs prises deux à deux). L'expression variable chétotaxique recouvrira les entités proprement dites et leurs combinaisons. Une **variable chétotaxique** peut être constante ou non. Les différentes valeurs pouvant être prises par une variable sont appelées des classes.

III. LES FONCTIONS DE DESCRIPTION

Soit un ensemble de p variables chétotaxiques, chacune contenant q observations ; une **fonction de description** f reliera chacune des variables ayant une image par f à un ensemble d'un ou plusieurs nombres appelés **descripteur chétotaxique**. Les descripteurs chétotaxiques permettent de rendre compte « de ce qu'il y a » dans les variables. L'ensemble des variables ayant une image par f constitue le domaine de définition de f . Nous distinguerons plusieurs fonctions.

1. La fonction intervalle de premier type

Par définition, l'image d'une variable chétotaxique par une fonction intervalle de premier type de niveau $p\%$ sera le plus petit intervalle des valeurs prises par la variable qui, après sélection et élimination des classes marginales (le plus symétriquement possible), comprend au moins $p\%$ des observations. Le descripteur chétotaxique est donc dans ce cas un couple de nombres correspondant aux bornes de l'intervalle en question.

La fonction intervalle de premier type sera utilisée dans la description des espèces. Cette fonction peut être appliquée à des variables, dont chacune intéressant une portion d'antenne est formée par la somme des entités chétotaxiques des huit génératrices. Il s'agit donc du nombre total de soies portées par une région de l'antenne.

La région en question sera généralement le tronçon B en entier, ou la zone BM. Le niveau retenu sera 80 %, ou bien 100 % lorsque la variable chétotaxique est constante ou quasi constante. L'usage nous a montré que le niveau de 80 % donne les meilleurs résultats pour décrire la variabilité du revêtement. La fonction donnant l'intervalle des classes comprenant au moins 80 % des observations sera écrite $n80\%$. L'expression $n80\%(BM)$ désignera le descripteur chétotaxique qui, pour une espèce donnée, correspond à l'intervalle contenant au moins 80 % des observations relatives à la variable « nombre total de soies G sur BM ».

Un exemple est donné sur la figure 3 qui présente l'histogramme du nombre de soies sur BM chez *Bourletiella hortensis* ; pour cette espèce : $n80\%(BM) = 19-26$ et $n80\%(B) = 32-39$.

Dans le cas où il n'y a aucune variabilité (nous ne considérerons pas de rares anomalies), l'intervalle des totaux de soies sur la zone BM se réduit à un seul nombre : $n80\%(BM) = n100\%(BM)$, que nous écrirons plus succinctement $n(BM)$. Ainsi chez *Lipothrix lubbocki* tout le tronçon B a une chétotaxie définie pour laquelle $n(BM) = 6$ et $n(B) = 14$.

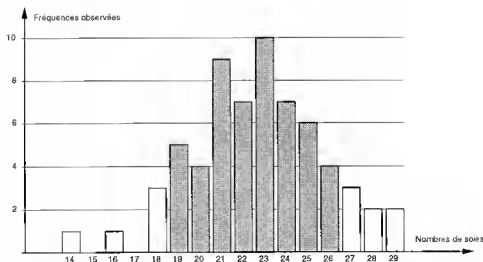


FIG. 3. — Histogramme du nombre de soies portées par la zone BM chez *Bourletella hortensis*; les colonnes en gris correspondent aux valeurs retenues pour la fonction intervalle de premier type au niveau 80 %. On a : $n80\%(BM) = 19-26$.

Enfin remarquons que, dans le cas où il n'y a pas de soie variable sur BB et sur BA, nous pouvons aisément déduire $n80\%(B)$ de $n80\%(BM)$ en ajoutant aux deux valeurs limites de l'intervalle la somme du nombre de soies sur BA avec le nombre de soies sur BB. C'est le cas de *B. hortensis*, chez qui cette somme est $8 + 5 = 13$; en ajoutant 13 aux valeurs limites $n80\%(BM)$ (19 et 26) nous trouvons directement les valeurs limites de $n80\%(B)$ (32 et 39).

L'intérêt du calcul de ces descripteurs se situe dans la comparaison de l'intensité de la pilosité du tronçon basal entre taxons, ou même entre sexes comme chez *Deuterosminthurus sulphureus* (Koch, 1840) puisque cette espèce montre un dimorphisme sexuel concernant le nombre de soies portées par BM. Ce dimorphisme sexuel a déjà été mis en évidence par BRETFELD (1990) chez *Heterosminthurus insignis* (Reuter, 1876) et *Deuterosminthurus bicinctus* (Koch, 1840).

2. La fonction intervalle de deuxième type

Une fonction intervalle de deuxième type donne un résultat plus nuancé : outre l'élimination des classes marginales de faible effectif, elle distingue des classes majeures et des classes mineures. La seule fonction intervalle de deuxième type qui nous intéresse est définie sur des variables qui sont des entités chétotaxiques plurielles (il s'agit en fait des génératrices d'un ensemble chétotaxique structuré non défini).

La loi de distribution dans les diverses classes présente généralement un maximum et des valeurs décroissantes autour de celui-là. La description que nous présentons permet un

resserrement des valeurs et une indication de l'importance relative entre les effectifs des classes conservées. Nous proposons la démarche suivante :

Premier temps : élimination des valeurs marginales à droite et à gauche de l'histogramme. La valeur marginale est ici définie comme la réunion, la plus grande possible, des classes d'une marge soumise aux deux conditions suivantes : le pourcentage de chaque classe doit être inférieur ou égal à 10 % et la somme des pourcentages ne dépasse pas 15 %.

Deuxième temps : distinction de classes majeures et de classes mineures. Chaque classe dont l'effectif est inférieur de moitié à celui de la classe de plus grand effectif sera considérée comme classe mineure, les autres seront les classes majeures. Lorsqu'il n'existe pas de classe mineure, toutes les classes sont majeures.

Troisième temps : écriture du descripteur chétotaxique. Les valeurs retenues sont classées par ordre croissant avec les classes mineures entre parenthèses.

Cette fonction qui donne une description de la variabilité dans les génératrices sera écrite *v*. Par exemple pour *Ge* sur *BM* d'*Arrhopalites terricola* Gisin, 1958, nous avons les observations suivantes : nombres de soies : 1, 2, 3 ; fréquences observées : 1, 33, 7.

La fonction intervalle de deuxième type permettra de présenter de manière condensée le nombre de soies présentes sur cette portion de génératrice : $v(Ge \cap BM) = 2;(3)$.

3. La fonction occurrence

Cette fonction est définie pour des entités chétotaxiques singulières. Pour une soie donnée, nous calculons d'abord sa probabilité de présence. Puis nous distinguons deux cas :

- 1) ou bien la soie en question appartient à un groupe de soies dont rares sont celles qui sont variables (c'est le cas de la quasi-totalité des soies *G*) ;
- 2) ou bien la soie en question appartient à un groupe de soies pour lesquelles la variabilité est courante (c'est le cas des soies *H*).

Dans le premier cas : si la probabilité de présence est strictement inférieure à 0,2, l'occurrence est 0 ; si cette probabilité est comprise entre 0,2 inclus et 0,8 inclus, l'occurrence est égale à cette probabilité arrondie au premier chiffre après la virgule ; si cette probabilité est strictement supérieure à 0,8, l'occurrence a pour valeur 1.

Dans le deuxième cas, l'occurrence est égale à la probabilité de présence donnée avec un chiffre significatif.

Cette démarche nécessite quelques explications. Les soies *G* de l'antenne ainsi que des autres appendices sont assez peu variables. Et lorsqu'elles le sont (il s'agit surtout de soies secondaires), leur variabilité est faible ; c'est-à-dire que leur probabilité de présence est voisine de 1 ou de 0. D'où l'idée qu'une soie vue sur plus de 4/5 des observations puisse être considérée comme présente et qu'une soie se trouvant sur moins de 1/5 des observations puisse être regardée comme absente. Dans l'absolu ces phanères sont variables, néanmoins vouloir les considérer comme tels conduit à une multiplication des cas de variabilité, ce qui complique beaucoup les descriptions de taxons ou la manipulation des données lors d'une analyse ontophylogénétique : les raisons pratiques ne sont pas étrangères à nos choix. Ainsi, les termes

présent et absent ont ici les sens de « souvent présent » (sur les observations qui ont été menées) et de « souvent absent ».

On peut alors se demander quelle est la probabilité d'erreur lorsqu'on affirme qu'une soie est présente, ou bien qu'elle est absente, alors qu'elle est franchement variable. Prenons un exemple simple : soit un phanère présent dans une population au taux de 0,5 — donc vraiment variable — nous observons 4 individus, nous menons donc 8 observations sur des organes qui sont paires. La probabilité que nous donnions la valeur 1 à l'occurrence se réalise pour les deux événements suivants : 8 soies observées et 7 soies observées ; cette probabilité vaut 3,5 %. La probabilité de donner la valeur 0 à l'occurrence vaut aussi 3,5 %. Ainsi, la probabilité de déclarer cette soie constante est de 7 % ; eu égard au faible nombre d'observations (seulement 8 !), ce risque paraît assez faible.

Pour les soies H nous préférons garder toute l'information sur l'intervalle 0-1. En effet, ces soies sont souvent variables et comme nous le verrons elles présentent un comportement collectif, ce qui nous amènera à faire la somme des occurrences des soies portées par une intergénératrice.

La fonction occurrence est notée *oc*. Par exemple, chez *Capraínea marginata* (Schött, 1893) $oc(AQ)BA\pi = 0,4$. Pour la soie en question, dans un souci de concision, nous écrirons aussi $oc = 0,4$. Cette remarque peut être étendue aux autres fonctions : en fait cela revient à confondre dans l'écriture la fonction de description et le descripteur calculé pour une soie.

4. Les fonctions statistiques

Nous emploierons plusieurs fonctions statistiques usuelles, notamment la moyenne (notée *m*), rarement l'écart-type. Remarquons que pour des entités chétotaxiques singulières, moyenne et probabilité de présence sont équivalentes.

Jusqu'à présent nous avons décrit des fonctions d'une variable chétotaxique. Nous pouvons aussi utiliser des fonctions de plusieurs variables chétotaxiques. Ces fonctions permettent d'attribuer des coefficients de dépendance à des groupe de variables. Parmi les coefficients pouvant être employés citons le χ^2 , le coefficient de contingence Φ de PEARSON, les coefficients de corrélation totale, partielle ou multiple, etc. Chacun de ces coefficients pourra être testé, ce qui permet de conclure avec un risque d'erreur connu (et donc de manière objective) sur la réalité de l'existence d'une liaison statistique.

IV. L'INTERRELATION CHÉTOTAXIQUE

Sur le quatrième segment antennaire de *Sminthurus viridis* nous avons pu mettre en évidence les deux phénomènes suivants (NAYROLLES, 1991a) :

— de l'apex vers la base du tronçon M existe sur chaque intergénératrice un gradient décroissant des probabilités de présence des soies de chaque verticille ;

— dans ce gradient, les soies d'une même intergénératrice situées sur deux verticilles successifs montrent dans leur présence une corrélation négative très nette. En d'autres termes,

pour une même intergénératrice, le fait de savoir qu'une soie existe sur le verticille n réduit les chances de voir une soie sur les verticilles $n-1$ et $n+1$. Ce phénomène a été appelé une interrelation chétotaxique.

Ici il s'agit d'une interrelation chétotaxique négative. Elle se mesure à l'aide d'un « outil » statistique au sein d'une espèce : nous qualifierons cette interrelation d'intraspécifique. Pour son calcul, les fonctions statistiques les plus utilisées seront le χ^2 , le coefficient de contingence Φ de PEARSON, le coefficient de corrélation r de BRAVAIS-PEARSON, etc.

V. LA COMPENSATION

Sur une population toulousaine de *Deuterosminthurus sulphureus*, pour les soies de Hppe, sur les cinq verticilles du tronçon M, nous avons les occurrences suivantes : $oc(M1ppe) = 1$; $oc(M2ppe) = 1$; $oc(M3ppe) = 0,8$; $oc(M4ppe) = 0,3$ et $oc(M5ppe) = 0$. Nous avons pu montrer que M3ppe et M4ppe sont en interrelation chétotaxique négative (test très significatif). La proportion des observations pour lesquelles le nombre de soies sur Hppe est de 3 est donc supérieure à la probabilité d'observer ce nombre en cas d'indépendance entre les deux soies M3ppe et M4ppe. Dans une population aveyronnaise de *D. sulphureus*, les occurrences des soies de Hppe sur le tronçon M, de l'apex vers la base, sont les suivantes : 1 ; 1 ; 1 ; 0 ; 0. Le nombre de soies sur Hppe est donc de 3.

Ainsi, il semble que « l'important » n'est pas l'occurrence de chaque soie mais la somme des occurrences ; cette valeur est la plus conservée, la plus caractéristique pour l'espèce. Dans la population de Toulouse l'absence de soie sur M3 est compensée par la présence d'une soie sur M4 et réciproquement ; ce phénomène est une **compensation**.

La compensation a pour effet de resserrer le spectre des fréquences du nombre de soies d'un ensemble chétotaxique. A cet égard le tableau I est très démonstratif. Ce tableau présente

TABLEAU I. — Fréquences des nombres de soies de Heae sur M chez *Sminthurus viridis* pour des individus ayant 14 verticilles sur M (d'après NAYROLLES, 1992b). Les fréquences théoriques correspondent à l'indépendance des soies de Hppe entre elles. $\chi^2 = 15,03$; DL = 4 ; χ^2 significatif au seuil de 1 %.

Nombres de soies de Heae sur M	≤ 7	8	9	10	$11 \leq$
Fréquences observées	1	12	27	10	0
Fréquences théoriques	5,6	12,3	16,4	11,5	4,2

TABLEAU II. — Tableau de contingence : fréquences observées des nombres de soies du tronçon M sur Heae et Hppe chez des individus de *S. viridis* ayant 14 verticilles sur M (d'après NAYROLLES, 1992b). $\chi^2 = 15,52$; DL = 2 ; χ^2 significatif au seuil de 1 %.

		Nombres de soies de Heae sur M		
		7 & 8	9	10
	9 & 10	1	13	9
Nombres de soies de Hppe sur M	11 & 12	12	14	1

le phénomène de compensation sur Heae chez *S. viridis* pour des individus ayant 14 verticilles sur M. Les résultats pour Hppe et pour des individus avec 15 verticilles sur M sont similaires.

L'interrelation chétotaxique négative entre soies voisines de même intergénéatrice observée chez *S. viridis* est très probablement liée au phénomène de compensation.

Alors que les soies de Heae prises isolément ne présentent pas d'interrelation chétotaxique significative avec celles de Hppe, considérés collectivement ces deux ensembles montrent une liaison. Le tableau II la met en évidence. La compensation présentée dans ce tableau concerne le cas d'individus à 14 verticilles sur M ; le phénomène se retrouve bien sûr pour des animaux ayant 15 verticilles sur M.

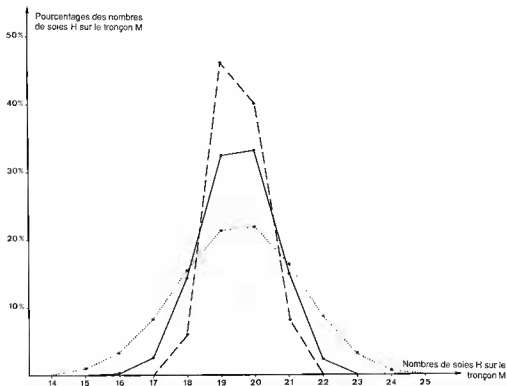


FIG. 4. — Pourcentages des nombres de soies H sur le tronçon M chez *S. viridis* pour les individus à 14 verticilles sur M (d'après NAYROLLES, 1992b). Histogrammes des valeurs observées et des valeurs calculées selon deux modèles.

Trait discontinu = les valeurs observées ; trait continu = le premier modèle : compensation au sein de Heae d'une part, de Hppe d'autre part, et aucune compensation entre Heae et Hppe ; pointillés = le deuxième modèle : aucune compensation.

Soulignons que les seules intergénératrices présentes chez *S. viridis* étant Heae et Hppe, le nombre de soies H est égal à la somme des nombres de soies sur Heae et sur Hppe.

Nous avons dressé, chez *S. viridis*, pour des individus possédant 14 verticilles sur M (50 observations), le graphique (fig. 4) des pourcentages observés des différents nombres de soies H sur M, et nous l'avons confronté à deux modèles théoriques.

Le premier modèle est calculé en prenant les pourcentages des nombres de soies de Heae sur M d'une part, et de Hppe sur M d'autre part, tels qu'ils furent observés. A partir de ces nombres on peut connaître les pourcentages des nombres de soies H sur M tels qu'ils se présenteraient en l'absence de compensation entre intergénératrices (indépendance statistique entre Heae et Hppe). Ce modèle ne prend donc en compte que la compensation au sein des intergénératrices.

Dans le deuxième modèle, ont été calculés les pourcentages des nombres de soies de Heae sur M à partir des probabilités de présence des soies de l'intergénératrice en question sur les verticilles successifs de M, ces soies étant considérées comme statistiquement indépendantes. La même opération est faite pour Hppe. Puis les calculs pour obtenir les pourcentages des nombres de soies H sur M sont menés comme dans le premier modèle. Le plan que nous venons de décrire permet ainsi d'annuler tout phénomène de compensation (tant au sein de Heae, que de Hppe, et qu'entre Heae et Hppe).

Au vu de la figure 4, il est évident que le resserrement des valeurs prises par la variable « nombres de soies H sur M » croît lorsque la compensation augmente.

Nous avons pu vérifier l'existence du phénomène de compensation sur plusieurs espèces de Sminthuridae et de Bourletiellidae. D'après nos observations, nous pouvons dire que dans ces deux familles, pour les intergénératrices Heae et Hppe, les occurrences des soies H considérées isolément sont de faible valeur informative. Dans un premier temps, il conviendra de garder comme grandeur devant servir de comparaison entre taxons le nombre de soies sur Heae d'une part et sur Hppe d'autre part. Nous exposerons dans un prochain travail pourquoi nous considérerons comme suffisante la variable correspondant au nombre total de soies sur Heae et sur Hppe.

CONCLUSION

Essayons tout d'abord de faire une analyse de nos points de vue sur le traitement mathématique à apporter aux soies variables.

Dans le cas d'une soie G idionymique, sa variabilité sera donnée par un nombre : l'occurrence (oc). Le plus souvent, de telles soies ne montrent pas de comportement collectif bien distinct (à l'inverse des soies H). Aussi nous assimilerons une soie fréquente (plus de 4 observations sur 5) à une soie présente (et $oc = 1$) ; inversement une soie rare (moins d'une observation sur 5) sera considérée comme absente ($oc = 0$) ; entre les deux, la valeur de l'occurrence s'échelonnara de 0,2 à 0,8 par pas de 0,1.

L'intérêt de borner l'intervalle de variabilité à 0,2 — 0,8 se situe dans la limitation du nombre de phanères variables précisé dans chaque description d'espèce.

La notion d'occurrence correspond à un traitement qui se veut général en face de ce type de variabilité.

Dans la problématique de la variabilité des phanères, nous l'avons déjà souligné, seul GRANDJEAN entreprit des travaux d'importance.

Face à quasiment le même type de variabilité que nous venons d'exposer, la réponse de GRANDJEAN (1942) fut complètement différente de la nôtre. Voici rapidement expliquée la méthode qu'il proposa : soit p la probabilité de présence d'une soie ; si p est significativement supérieure à 0,5, alors la soie est dite probable pour la stase (le stade de développement) en question ; si p est significativement inférieure à 0,5, alors la soie est non probable ; et si p est voisine de 0,5, la soie est dite quasi probable. Néanmoins, GRANDJEAN objecte lui-même : « L'expression probable, d'après sa définition, a ici l'inconvénient de confondre toutes les fréquences pourvu qu'elles soient supérieures à 0,5. » Cet inconvénient n'existe pas dans le cas de l'occurrence. Nous verrons dans de prochaines études l'intérêt de l'utiliser.

Nous avons bien conscience de la forte imprécision attachée à l'occurrence que nous attribuons à un phanère ; en effet, pour un faible nombre d'observations l'intervalle de confiance d'une proportion est très grand. Néanmoins le temps nécessaire à la préparation des échantillons et à leur étude limite le nombre d'observations réalisables.

Ce point de discussion établi, tentons maintenant de rassembler ce qu'il y a lieu de retenir.

Tout d'abord à travers ce travail nous avons voulu montrer qu'un groupe de soies peut être divisé jusqu'à des éléments qui ne sont plus décomposables. Ceux-là constituent les objets d'étude du taxonomiste : ce sont les entités chétotaxiques. L'acquisition de ces objets correspond à une démarche réductrice puisque nous les définissons sur chaque espèce, à chaque stade et éventuellement sur chaque sexe. Le cheminement inverse nous amène à considérer des ensembles chétotaxiques « globaux » pour un groupe.

Les entités chétotaxiques sont des vecteurs d'observations et correspondent à des données brutes. Pour en « tirer quelque chose », nous leur ferons subir deux types de « traitements ». Un premier qui, en les combinant entre elles, formera des variables chétotaxiques élaborées. Un deuxième qui consistera à tirer les renseignements contenus dans ces variables. Cela se fera à l'aide d'une fonction reliant un vecteur d'observations à une description qui, d'une façon très brève, rend compte de la majeure partie de l'information des données.

Nous avons aussi vu que les soies peuvent présenter des liaisons statistiques entre elles. Le traitement d'un tel problème présente beaucoup d'analogie avec ce que nous venons de voir. Ainsi une fonction statistique donnera une description de la liaison existant entre deux vecteurs d'observations.

Enfin les liaisons entre soies peuvent être d'ordre collectif ; ce qui est illustré par la compensation. Ce phénomène aboutit à un resserrement de l'intervalle des valeurs prises par une variable du type « nombre total de soies » d'un ensemble chétotaxique (ce qui est fort intéressant pour le taxonomiste).

En conclusion, nous avons posé ici le problème des objets dans l'étude du revêtement : comment les identifier, l'information qu'on doit pouvoir en tirer, la mise en évidence de liaisons pouvant les unir.

Remerciements

Il me tient à cœur d'exprimer ma reconnaissance à MM. L. BONNET et J.M. BETSCH qui ont bien voulu lire mon manuscrit et me faire part de leurs remarques.

RÉFÉRENCES

- ANDRÉ, H. M., 1986. — The stases in Collembola : a case study. *In* : R. DALLAI (ed.), Second International Seminar on Apterygota. Università di Siena, Italy : 301-305.
- 1988. — Age-dependent evolution : from theory to practice. *In* : C. J. HUMPHRIES (ed.), Ontogeny and Systematics. Columbia University Press, New York, 137-187.
- 1989. — Ontogenetic trajectories in arthropods. *In* : H. M. ANDRÉ & J.-Cl. LIONS (eds), L'ontogenèse et le Concept de Stase chez les Arthropodes / Ontogeny and Concept of Stase in Arthropods. AGAR Publishers, Wavre, Belgium : 83-90.
- BRETFELD, G., 1990. — Chaetotaxy of four Species of the Genera *Heterosminthurus*, *Bourletiella*, *Deuterosminthurus* and *Prorastriopes* (Insecta, Collembola, Symphypleona). *Zool. Jb., Syst.*, **117**, 441-489.
- GRANDJEAN, F., 1938. — Sur l'ontogenèse des Acariens. *C. r. Acad. Sci., Paris*, **206** : 146-150.
- 1939. — La répartition asymétrique des organes aléatoires. *C. r. Acad. Sci., Paris*, **208** : 861-864.
- 1942. — Les méthodes pour établir les listes de priorité et la concordance de leurs résultats. *C. r. Acad. Sci., Paris*, **214** : 729-733.
- 1949a. — Sur les rapports théoriques entre écarts et mutations. *C. r. Acad. Sci., Paris*, **228** : 1675-1678.
- 1949b. — Au sujet des variations individuelles et des polygones de fréquence. *C. r. Acad. Sci., Paris*, **229** : 801-804.
- NAYROLLES, P., 1991a. — La chétotaxie antennaire des Collemboles Symphypléones. *Trav. Lab. Écobiol. Arthr. édaph. Toulouse*, **6** (3) : 1-94.
- 1991b. — Contribution à l'étude ontophylogénétique et écologique des Collemboles Symphypléones épigés. Thèse de l'Université P. Sabatier, Toulouse : 1-329.
- 1992. — Aspects structuraux de la chétotaxie appendiculaire des Collemboles Symphypléones. *Nouv. Revue Ent. (N.S.)*, **9** (4) : 345-356.